Coordination of Programs on Domestic Animal Genomics: The Federal Framework

Progress Report

June 2004

National Science and Technology Council
Committee on Science
Interagency Working Group on Domestic Animal Genomics

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUN 2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **Coordination of Programs on Domestic Animal Genomics: The Federal Framework**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Science and Technology Council,Executive Office of the President,725 17th Street Room 5228,Washington,DC,20502**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**37**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

About the National Science and Technology Council

The National Science and Technology Council (NSTC), a cabinet-level council, is the principal means for the President to coordinate science and technology policies across the Federal Government. NSTC acts as a "virtual" agency for science and technology to coordinate the diverse parts of the Federal research and development enterprise.

An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems, strengthening fundamental research, securing the homeland, and strengthening the economy. This council prepares research and development strategies that are coordinated across Federal agencies to form an investment package that is aimed at accomplishing multiple national goals.

Additional information regarding the NSTC can be obtained from the NSTC website, http://www.ostp.gov/NSTC/html/NSTC_Home.html.

Coordination of Programs on Domestic Animal Genomics: The Federal Framework

Progress Report
June 2004

National Science and Technology Council

Committee on Science

Interagency Working Group on Domestic Animal Genomics

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
WASHINGTON, D.C. 20502

September 23, 2004

Dear Colleague:

This report provides an update on the five-year plan for Federal agencies dealing with genomics activities related to domesticated animals. This plan was developed by the NSTC Committee on Science's Interagency Working Group (IWG) on Domestic Animal Genomics in 2003. The IWG was originally chartered in March 2002, to raise awareness of the importance of domesticated animal species and to address and coordinate Federal programs in domestic animal genomics. The IWG consulted both scientists and industry representatives in developing the five-year "road map" plan.

Significant milestones have been achieved in the past few years in genomics research on humans, research model animals, plants, and microorganisms, both domestically and internationally. The extension of these efforts to include the domestic animal species offers the opportunity to elucidate the mechanisms underlying human disease via more complete understanding of evolutionary relationships between species leading to the development of new therapeutics. Additionally, domestic animal genomics offers the opportunity to improve animal health, increase food quality, safety and nutritional value, improve human nutrition, and increase production efficiency of animal agriculture.

Judging from the progress made in the past year through activities described in this report, there is every indication that continued significant advances will be achieved in the coming year, particularly in completing critical infrastructure in the areas of structural genomics and bioinformatics. The IWG will continue to coordinate the activities contributing to the five year plan to ensure that U.S. efforts in domestic animal genomics benefit from interagency support and cooperation, keeping U.S. scientists at the forefront of animal biology and its application to solving global problems in public health, agriculture, energy, and environmental protection.

Sincerely,

Kathie L. Olsen
Co-Chair
Committee on Science

Arden Bement
Co-Chair
Committee on Science

Elias A. Zerhouni
Co-Chair
Committee on Science

Interagency Working Group on Domestic Animal Genomics
Committee on Science
National Science and Technology Council

Chair

Joseph J. Jen
Under Secretary
Research, Education and Economics
U.S. Department of Agriculture

Members

Robert B. Bertram
Chief
International Research and Biotechnology Division
Office of Environment and Science Policy
U.S. Agency for International Development

Adrienne Erbach
Chief
Agriculture Branch
Natural Resource Programs
Office of Management and Budget
Executive Office of the President

Aristides Patrinos
Associate Director
Office of Biological and Environmental Research
U.S. Department of Energy

William Zamer
Program Director
Ecological and Evolutionary Physiology
Division of Integrative Biology and Neuroscience
National Science Foundation

Lester Crawford
Deputy Commissioner
Food and Drug Administration

Cliff Gabriel
Deputy Associate Director for Science
Office of Science and Technology Policy
Executive Office of the President

Jane Peterson
Associate Director
Extramural Research
Program Director
Comparative Genomics Program
National Institutes of Health

Ronnie D. Green (Executive Secretary)
National Program Leader
Food Animal Production
Agricultural Research Service
U.S. Department of Agriculture

Associates

Rodney J. Brown
Deputy Under Secretary
Research, Education and Economics
U.S. Department of Agriculture

Leland Ellis
Senior Liaison for Agriculture
Science and Technology
U.S. Department of Homeland Security

Muquarrab A. Qureshi
National Program Leader
Animal Genetics
Cooperative State Research, Education and Extension Service
U. S. Department of Agriculture

Daniel W. Drell
Biologist
Human/Microbial Genome Program
Office of Biological and Environmental Research
U.S. Department of Energy

Marvin E. Frazier
Director
Life Sciences Division
Office of Biological and Environmental Research
U.S. Department of Energy

Table of Contents

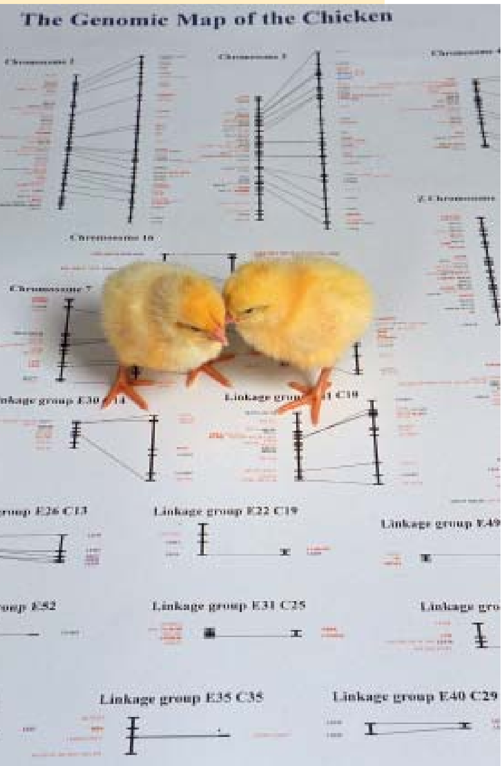
Coordination of Programs on Domestic Animal Genomics: The Federal Framework

Progress Report

I. Executive Summary

The Interagency Working Group (IWG) on Domestic Animal Genomics was chartered in 2002 with the charge of enhancing interagency communication and awareness of the importance of agricultural and companion animal species, increasing leverage of investments across government agencies, and positioning of agriculture as a critical element of the Federal genomics programs. The Committee on Science provides oversight over the IWG activities. The membership of the IWG consists of representatives from the Department of Agriculture (USDA), Department of Energy (DOE), Food and Drug Administration (FDA), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), Office of Management and Budget (OMB), and U.S. Agency for International Development (USAID). The IWG subsequently identified the following broad strategic goals:

- Bring into place the programmatic elements needed to advance the study and understanding of domesticated animal genomes, including large-scale DNA sequencing; functional characterization of expressed genes (functional genomics); tools for data storage, analysis and visualization (bioinformatics); and study of similarities among genomes (comparative genomics).
- Leverage the national infrastructure for large-scale DNA sequencing that has been established for the Human Genome Project and other vertebrate and model organism genomes.
- Advance and utilize the enabling tools and infrastructure of functional genomics and bioinformatics to enhance the understanding not only of basic science and disease mechanisms, but also to address critical agricultural missions, including animal health and well-being, food safety, and human nutrition.
- Ensure that genomics data are freely available in the public domain and genomics reagents and resources are available to the public.
- Increase the training opportunities for genomics and bioinformatics at all levels of education.
- Coordinate and encourage international cooperation to achieve these goals.



In September 2003, the IWG released its initial plan entitled "Coordination of Programs on Domestic Animal Genomics: A Federal Framework" in which large-scale sequencing, data management and bioinformatics, and functional genomics were identified as the *specific goals to be achieved in fiscal years 2003 to 2007* (see http://www.ostp.gov/NSTC/html Animal_Genome%7EWEB.pdf).

In this report, progress in the first year of the effort is presented. Highlights include:

- Completion of bovine and chicken BAC maps with the same nearing completion for swine.
- Progress on producing integrated physical and genetic maps for bovine, swine, and chicken.
- Completion of the first draft assembly of the honeybee and chicken genome sequences.
- Significant progress in completing the first assembly of the dog genome sequence.
- Successful launching of the bovine genome sequencing project.
- Activities of the International Swine Genome Sequencing Consortium in securing support to launch the porcine genome sequencing project.
- Successful application of genomics technology in addressing the US BSE cattle situation.

Also reported are plans for the coming year for the IWG, including:

- Development of a plan for bioinformatics and functional genomics tool development in agricultural animal research.
- Development of a broad-based functional genomics program in infectious animal disease.
- Expanding the structural genomics infrastructure to include additional species of interest.
- Consideration of genomics technology for animal traceability systems.

In the next year, all agencies participating in the IWG plan to continue support of domestic animal genomics research based on the five-year plan. The IWG and participating agencies recognize the rapid pace of developments in animal genomics and that there will likely be new opportunities requiring IWG attention as they become evident. The members of the IWG look forward to continuing to work to effectively coordinate and facilitate federal efforts in this important area.

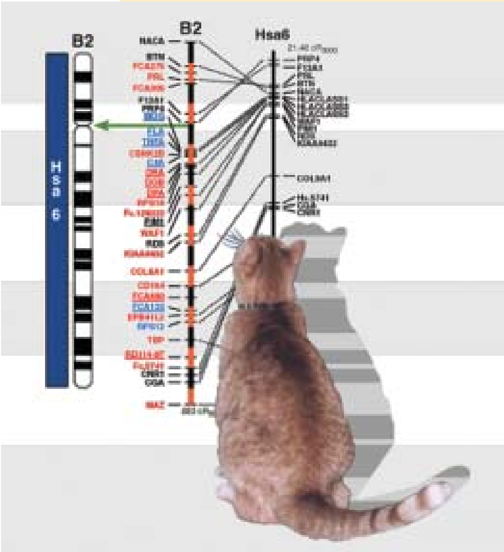
II. Introduction

Domesticated animals have played a key role in human society through their use as livestock, medical research models, and human companions. The completion of the human genome DNA sequence in April 2003 marked an important milestone in scientific knowledge. The research arising from it, called genomics, provides a new opportunity to improve and understand domesticated animals. Once a genome sequence is in hand, scientists use it to identify new genes, discover and understand regulatory elements, and study individual genes, their functional products on a molecular level, and their interactions with other genes. Additionally, the sequence can be used to study the evolution of genomes that will aid in understanding how molecular mechanisms arose.

In the United States, livestock and companion animals are the backbone of billion-dollar industries. Genomic studies of domesticated animals will yield an understanding of the genetics and expression of genes important to improving these species for greater food yields and improved animal health and human nutrition. Even more importantly, comparative genomics will make significant contributions to biomedical research and ultimately improved human health. Understanding human biology based on insights from a diversity of species will lead to advances in biomedical research and it is anticipated that it will accelerate the development of new human and animal pharmaceuticals.

The inclusion of domesticated animals in a Federal genomics program serves two primary purposes: 1) to increase the diversity of genomic sequence available for studies that by comparison with the human genome DNA sequence contribute to finding new genes and studying evolution at the genome level; and 2) to generate genetic data critical to efficient and sustainable animal agricultural production.

Genome sequencing projects are completed or underway for several animals (including the laboratory mouse, rat, rhesus macaque) that serve as models for studying human disease. Projects for sequencing the genomes of several domesticated animals (chicken, cattle, honeybee and dog) have now entered the genomic sequencing pipeline.



III. Background

The charge to the Interagency Working Group on Domestic Animal Genomics (IWG), which was established in 2002, was to enhance interagency communication and awareness of the importance of agricultural and companion animal species, to increase leverage of investments across government agencies, and to position agriculture as a critical element of the Federal genomics programs. The National Science and Technology Council's Committee on Science provides oversight over the IWG activities. The membership of the IWG consists of representatives from the Department of Agriculture (USDA), Department of Energy (DOE), Food and Drug Administration (FDA), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), Office of Management and Budget (OMB), and U.S. Agency for International Development (USAID).

Upon convening, the initial deliberations regarding the task set forth to the IWG led to an overarching statement of purpose:

> **The mission of the Interagency Working Group for Domestic Animal Genomics is to enhance communication and awareness of livestock and companion animal species of importance to the food and agriculture system; leverage Federal investments in large-scale genome sequencing and genome analysis across government agencies; position the food and agriculture system as a critical element of the national genomics program; enhance dialogue and cooperation among Federal agencies, universities, and industry in the nation; and promote international cooperation on domestic animal genomics research.**

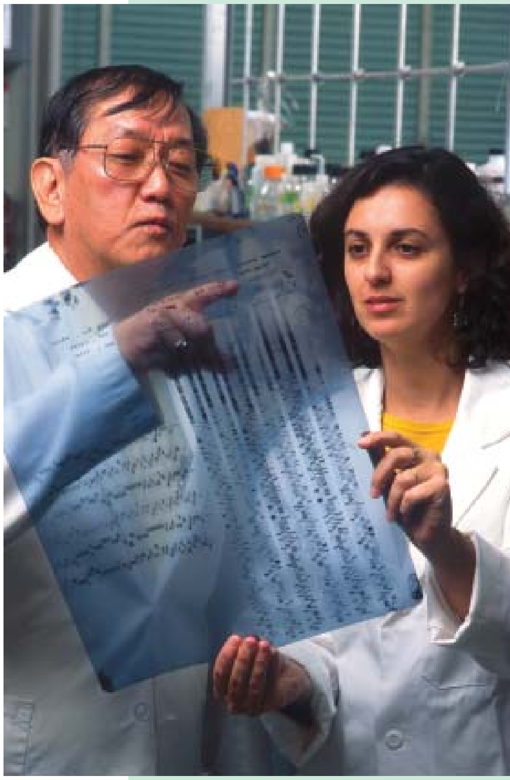
The IWG subsequently identified the following broad strategic goals:

- Bring into place the programmatic elements needed to advance the study and understanding of domesticated animal genomes, including large-scale DNA sequencing; functional characterization of expressed genes (functional genomics); tools for data storage, analysis and visualization (bioinformatics); and study of similarities among genomes of different species (comparative genomics).
- Leverage the national infrastructure for large-scale DNA sequencing that has been established for the Human Genome Project and other vertebrate and model organism genomes.

- Advance and utilize the enabling tools and infrastructure of functional genomics and bioinformatics to enhance the understanding not only of basic science and disease mechanisms, but also to address critical agricultural missions, including animal health and well-being, food safety, and human nutrition.
- Ensure that genomics data are freely available in the public domain and genomics reagents and resources are available to the public.
- Increase the training opportunities for genomics and bioinformatics at all levels of education.
- Coordinate and encourage international cooperation to achieve these goals.

The IWG determined that large-scale sequencing, data management and bioinformatics, and functional genomics are the *specific goals to be achieved in fiscal years 2003 to 2007,* including:

- Large-scale sequencing to produce draft genome sequences (8-fold sequence coverage) of honeybee, chicken, dog, cattle, swine, and cat.
- Data management and bioinformatics to specifically support agriculturally important species. This will support significant improvements in data management and analysis software, allow greater data accessibility and secure long-term maintenance, increase capabilities to deal with rapidly accumulating data complexity as databases include functional information, and provide more powerful tools to mine large genomes (individually and comparatively).
- Functional genomics to specifically investigate agriculturally important species. An increase in data for livestock genomes requires concomitant investment in functional genomics to support genome annotation, the study of gene regulation and expression, and species evolutionary relationships. Researchers will have access to individual genes, which can be cloned and characterized, the ability to scan entire genomes for specific functions, and be able to gain new insights from comparisons to other genomes. Quantitative trait loci will be accessible on a gene level, leading to studies that will elucidate characteristics such as food productivity and disease resistance and will lead to the ability to manipulate those traits more quickly and effectively.



IV. Progress Reported in the Past Year

Focus on Elucidating Genome Structure and Organization

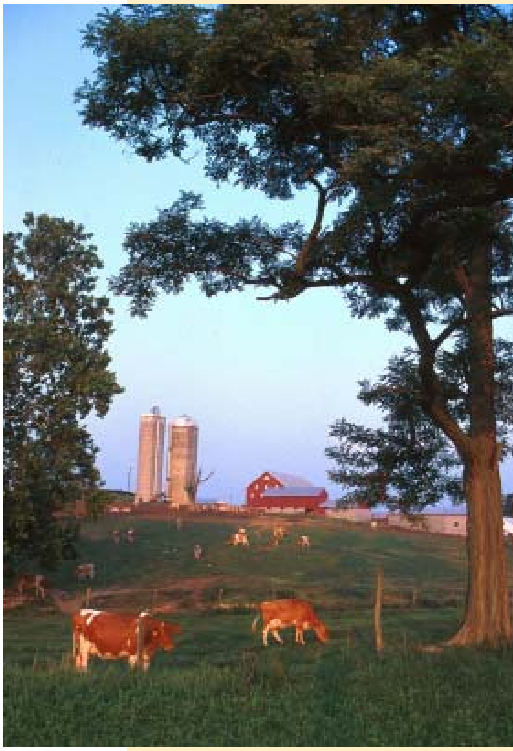
The progress reported herein builds upon an impressive foundation of agricultural and biological science. Research efforts, primarily over the past century, have provided a vast amount of phenotypic and genotypic information for a wide array of characters in livestock and companion animals. Long-term large population studies have been employed, particularly in the area of livestock production science, to develop technology to allow genetic evaluation and subsequent selection to be practiced on breeding populations in highly sophisticated planned breeding programs, especially for dairy and beef cattle, layer and broiler chickens, and swine. As DNA-based technologies were developed in the 1970s and 1980s, these well-characterized populations were invaluable in providing opportunities to develop genetic linkage maps to begin the search for genes affecting important traits. Highly saturated genetic maps were developed for all of the major domesticated animal species during the 1990s, allowing Quantitative Trait Loci (QTL) to be identified across many species and for many measures of performance. In parallel, genetic studies of the human were progressing with the development of genetic maps that were used to identify genes important in human disease. It was recognized in the late 1980s that progress in developing tools for agricultural or biomedical application could be greatly accelerated by producing high-resolution physical maps and ultimately the genome sequence of human and agricultural species. The genome sequence is needed in order to find all of the genes in an organism and begin to unravel how genes function and interact (i.e. functional genomics). Knowledge of gene interactions leads to further research in understanding proteins and their interactions and networks/pathways within the cell.

The challenge presented by the ambitious goals set for the human genome project stimulated rapid improvements in DNA sequencing strategies and technologies in the past decade. Following the successful completion of the human genome sequence and substantial progress in producing the sequence of the major model species used in biomedical research, there continued to be a need for sequencing of additional mammalian genomes, including domesticated animals, to be used as a tool to find all of the genes and regulatory elements encoded in the human genome. The approach used, called comparative genomics, compares and searches for similarities in DNA sequences from distantly related organisms. Less than half of these regions of sequence similarity contain genes known to be highly conserved throughout

the evolutionary tree, and it is postulated that the remaining regions may encode important regulatory elements. In the past two years, the National Human Genome Research Institute (NHGRI) of the National Institutes of Health included in its sequencing program the sequencing of genomes from species whose sequence was expected to help find all of the genes and identify regulatory elements in the human, mouse, and rat genomes through comparative sequencing and evolutionary biology approaches.

During 2002, proposals were submitted to NHGRI to develop draft-quality genome sequences for the chicken, cow, honeybee, pig, dog, and cat. All were designated to be "high priority" for sequencing by NHGRI supported sequencing centers (with the understanding that close evolutionary relationships between dog/cat and cow/pig would only allow justification for one of each pair of species to be sequenced in the NHGRI program) based upon the value of their sequence data to better understanding of the human genome thereby improving human health. Designation of high priority does not automatically mean that a project will be undertaken by the NHGRI sequencing program and in several cases, contributions from partners having a significant interest in obtaining the genomic sequence of particular organisms have been required. A primary emphasis of the IWG since its inception has been how to bring together the resources and partnerships to allow these large-scale sequencing projects to come to fruition for domestic animal species.

The following summaries highlight the major accomplishments during 2003 along the pathway to completion of the genomics infrastructure needed to advance the genomic studies of the chicken, cow, dog, honeybee, pig, and cat. The progress to date has occurred as a two-step process. The first step is the completion of infrastructure organizing the genetic (linkage) and physical {bacterial artificial chromosome (BAC) and radiation hybrid (RH)} maps for each of the important species. The resulting integrated genome maps then become the foundational "scaffolding" upon which whole genome sequencing is performed—resulting in the ultimate full sequence map of the genome.



Genetic and Physical Genome Maps

International Consortia Develop Bovine and Porcine BAC Maps

The availability of physical maps, and eventually the DNA sequence of the bovine and porcine genomes, will transform genetic research in these species. Such maps will greatly facilitate the identification of genes affecting production traits, disease susceptibility, animal health and product quality and nutritional

value. The elucidation of these genomes will also benefit human medicine by identifying novel genes and their function that are critical for human health. It is important that the genetic and physical maps, and the reagents used to build them, are freely accessible and in the public domain. The bovine and swine genomics research communities need additional laboratory resources to allow more efficient use of current human and mouse mapping information to solve production inefficiency and health problems and contribute to increasing the global competitiveness of these livestock industries.

Bovine BAC Map

An international effort (International Bovine BAC Mapping Consortium; IBBMC) was initiated to develop a physical BAC-based map of the bovine genome. Clones from three publicly available bovine BAC libraries were fingerprinted and end-sequenced to develop the bovine BAC map. The three BAC libraries were generated from three common breeds, Holstein, Hereford, and Angus. A total of 295,000 clones have been fingerprinted (15.8 X coverage) from these three libraries. Additional fingerprinting (100,000 clones) was performed on a fourth BAC library by INRA (L'Institut National de la Recherche Agronomique), France and is being end-sequenced (12,000 clones to date). Overall, a total of 160,000 clones have been sequenced with 140,000 clones having both ends sequenced. End sequencing information from an additional 100,000 clones by the INRA and international effort will be added and integrated in the near future. All sequence data are publicly available (www.ncbi.nih.gov/Genbank).

The construction of the integrated BAC map is in progress. The initial contig assembly generated 14,000 contigs with 2 or more clones per contig and 30,000 singletons. By utilizing the information in the human map and sequence, the contigs were merged reducing the number to 2,622 with 621 of these contigs containing 199,326 clones that are anchored to human chromosomes. A 2,000 marker radiation hybrid (RH) map has been generated that will be used to further merge contigs and anchor contigs to the human and bovine map. Publication of the bovine BAC map is planned for late 2004.

The BAC map was used as one of the justifications to sequence the bovine genome since this would be the first genome to be sequenced with a BAC map nearly completed prior to the start of sequencing. The sequencing of the bovine genome has now been funded and an element of the sequencing strategy will be to perform "light" sequencing of BAC clones across the genome. 19,000 clones have been selected from the Hereford library based upon the BAC map for this purpose. Additionally, selected BACs from regions of high research interest will be chosen for high-quality sequencing.

The participants of the IBBMC include USDA/ARS (Beltsville, Maryland and Clay Center, Nebraska), Texas A&M University, University of Illinois, The Institute for Genomic Research, Children's Hospital Oakland Research Institute, The University of Alberta and Alberta cattlemen (Alberta Livestock Genomics Initiative), INRA (France), AgResearch (New Zealand), CSIRO (Australia), EMBRAPA (Brazil), Roslin Institute (Scotland), British Columbia Cancer Agency Genome Sciences Centre (Canada), and the Alliance for Animal Genomics Research.

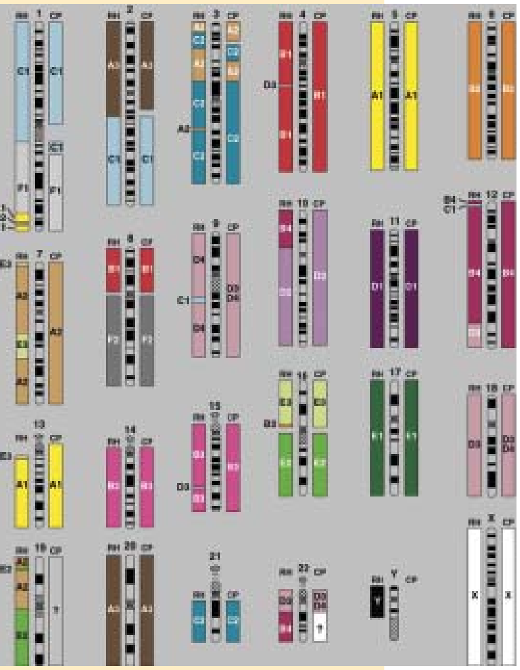
Swine BAC Map

The swine genomics research community organized a similar and parallel effort to that employed by the international bovine consortium to produce a porcine BAC map. The participants in the International Porcine BAC Mapping Consortium (IPBMC) include the University of Illinois, USDA/ARS U.S. Meat Animal Research Center, Roslin Intitute, Sanger Institute (U.K.), INRA, and the Korean Livestock Institute.

Clones from five available BAC libraries are being fingerprinted and end-sequenced in the swine effort (generated from Duroc, Korean, Meishan and Large White breeds). Thus far, a total of 300,300 clones have been fingerprinted with 225,000 clones having both ends sequenced. Additional clones are scheduled for end-sequencing in the spring of 2004. Similar to the model used by the bovine BAC-map consortium, the porcine map will be anchored to human chromosomes using a RH map currently being produced by integration of the 12,000-rad INRA and 7,000-rad INRA-University of Minnesota RH panels. All sequence data from this effort are being deposited in GenBank (www.ncbi.nih.gov/Genbank).

Construction of the integrated maps is to be conducted at the Sanger Institute during the second quarter of 2004. The IPBMC will meet at the Sanger Institute in June 2004 to review the initial map and to finalize plans for publication of the integrated BAC map. The first public presentation of the integrated map is slated for the International Society of Animal Genetics meetings in Tokyo in September 2004.

The plans of the IPBMC also include contribution of a minimum tiling path of clones from the Duroc library to be eventually used in developing the porcine whole genome sequence assembly (see subsequent article on page 20 of this report).



Integrated Maps Nearing Completion for Cattle, Pig, and Chicken

Integration of genetic linkage maps with physical maps and sequences is an essential component of efforts to link traits of agricultural interest (productivity, disease resistance, etc.) to the genes encoding alleles that determine the phenotype. This linkage requires genes or markers that are both polymorphic (capable of linkage mapping in populations) and easily assignable to physical maps. While there is growing interest in single nucleotide polymorphisms (SNPs), simple tandem repeats (STRs, also called “microsatellites”) have been most heavily employed in domestic animals, due to their widely polymorphic nature.

Statistical approaches to build integrated maps for livestock species can consolidate information from independent maps and capitalize on complementary resolution of physical and genetic data sets to resolve marker order. Genetic data used for linkage mapping can position markers separated by recombination along a chromosome, but lacks the resolution necessary to reliably order close markers. Physical mapping data sets, such as radiation hybrid (RH) panels, can reliably order close markers, but lack long-range resolution to order groups of widely separated markers. Simultaneously solved maps representing both linkage and RH data take advantage of each to order markers more reliably.

The first likelihood-based whole-genome bovine map integrating linkage and RH data was completed by USDA/ARS’s Roman L. Hruska U.S. Meat Animal Research Center (MARC) at Clay Center, Nebraska, and is now available to the International Bovine BAC Map Consortium. The map combines data from two radiation hybrid panels, the 3000-rad European Union (COMRAD) panel and the 5000-rad Texas A&M/University of Illinois panel, with linkage data from the MARC reference families. Over 6,300 distinct markers are represented, with 1,381 of these represented in the linkage and at least one RH data set. Alignment between 2,061 marker sequences and human genomic sequence provides the basis for a bovine-human comparative map. Overlap between marker and BAC end sequence alignments on the human genome connects 1,504 markers to the bovine BAC map, augmenting 574 markers physically assigned to BAC clones. Besides providing a bovine order to anchor BAC contig assembly and ultimately guide bovine sequence assembly, the integrated map has had a significant positive impact on the efficiency of the SNP discovery process to develop new markers targeted to chromosomal regions known to harbor loci influencing economically important traits. Further development of the integrated bovine map will incorporate BAC data, providing another level of resolution to refine marker order.

The swine whole-genome integrated map has been constructed in collaboration with France's INRA. More than 1,000 genetic markers from MARC were included in the analyses with over 3,000 markers ran across the INRA-University of Minnesota pig RH 7,000-rad panel. Approximately 750 markers were included in both data sets. The genetic map developed at MARC has been dramatically expanded recently and now contains over 3,000 typed loci. One thousand of the new markers are single nucleotide polymorphisms developed from EST sequence data generated in the pig. These markers are crucial to developing a high-resolution comparative map and permit integration of the genetic, RH and BAC maps in the near future.

Genetic linkage maps for the chicken expanded rapidly in the 1990s with the application of DNA-based marker systems. The consensus linkage map (Schmid et al., Cytogenet. Cell Genet. 90, 2000) covers about 4000 cM, including over 2000 markers. Early efforts to generate a physical map of the chicken genome led to the development of several BAC libraries (Crooijmans et al., Mammal Genome, 11, 2000; Lee et al., Anim. Gent. 34, 2003). These now have been used to generate genome-wide BAC contig maps that place chicken genes and markers in specific molecular locations (Ren et al., Genome Res., 13, 2003; Warren et al., manuscript in preparation). A first generation BAC map, partially integrated to the linkage map, is now available (Ren et al., Genome Research 13:2754) and a much-improved version will be published soon by the Washington University Genome Sequencing Center. This map will be integrated both to the linkage map and to the full genome draft sequence using the same approaches as previously described for cattle and swine.

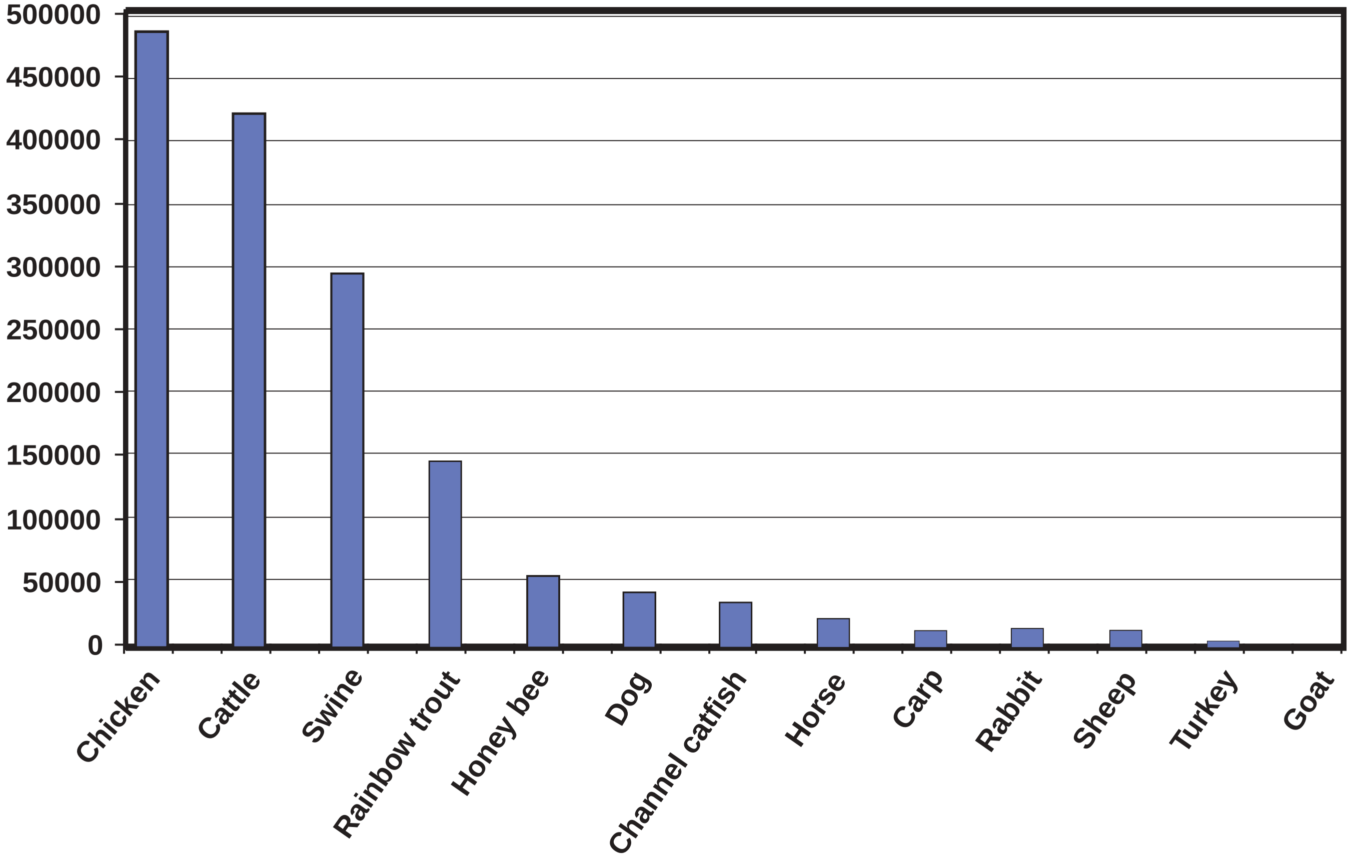
Given the draft level of the genome sequences now assembled for the chicken, in progress for the cow and proposed for the pig, map integration efforts will be important tools in assessing the quality of the sequence assemblies. The genome sequences will also provide marker sets (both SNP and STR) that will permit much higher resolution in future QTL mapping efforts for the purposes of mechanistic analyses and marker-assisted selection.

Domestic Animal EST Sequences in the Public Database Growing

The number of ESTs (Expressed Sequence Tags) in the dbEST database from domestic animals (http://www.ncbi.nlm.nih.gov/dbEST/) continues to increase at a rapid rate. As of February 2004, 1,530,000 ESTs had been deposited with the largest number coming from the chicken, cattle, and swine research communities.

Researchers are using the EST sequence information in dbEST in an increasing number of ways, particularly as mapping tools and for the development of gene expression profiling tools. The chicken is an excellent example of progress in this area. It is estimated that a total of over 600,000 chicken EST sequences have been determined (Burt and Pourquié, Science, 300, 2003) with more on the way. Chicken homologues to most mammalian genes can now be found in silico by searching the chicken EST and/or the non-redundant sequence databases (e.g., http://www.ncbi.nlm.nih.gov/blast). Brown et al. (Nature Rev. Genet. 4, 2003) reported that the chicken EST set already available showed strong matches to over 90% of the human disease genes in the "morbid map" of the Online Mendelian Inheritance in Man database (http://www.ncbi.nlm.nih.gov/Omim). Chicken EST contigs provide the basis for the construction of microarrays, used in global profiling of gene expression. Chicken gene microarrays have been used to analyze changes in mRNA levels in response to infection with Marek's Disease Virus (Liu et al., Animal Genet. 32, 2001) and Eimeria (Min et al., Appl. Microbiol. Biotechnol. 62, 2003) and in neoplasias induced by c-myc and c-myb over-expression (Neiman et al., Proc. Natl. Acad. Sci. U.S.A., 98, 2001, 2003). A 13,000 transcript chicken microarray (glass slide) is available from the Fred Hutchinson Cancer Research Center Genomics Facility, and more complete arrays using a variety of platforms are being planned.

Number of EST Entries in GenBank from Domestic Animals
(data from dbEST/GenBank/NCBI as of April 2004)



Genome Sequencing

Initial Draft of the Domestic Honeybee Genome Assembled

The honeybee (Apis meliferia) genome is the latest insect species, after the mosquito and flies, to be fully sequenced and made available to the public. Sequences from the honeybee will inform the medical sciences by improving research based on the Drosophila insect model, as well as providing information about behavior, development, and human immune reaction. The draft honeybee genome sequence is expected to provide insights into the impressive learning ability of these insects and may provide insights into learning and neurological diseases of humans. Additionally, human immune response to honeybee venom is an important medical problem in the United States. The honeybee sequence may also provide insights into important medical questions related to aging and reproduction. The honeybee genome sequence will also be used to address critical problems related to the agricultural importance of this species, ranging from disease resistance to nutrition, reproduction, and behavior. Finally, as a more distant relative of the three fly species whose genomes are available, honeybees provide a new reference point for comparative genomics across the insects, a group of worldwide importance for both human and ecosystem health. The request to NHGRI for sequencing of the honeybee can be found at http://www.genome.gov/10002154.

The honeybee genome sequence was produced from January to August of 2003, at the Baylor College of Medicine's Human Genome Sequencing Center and includes both short-insert or shotgun clones (2 million sequence reads) and end sequences from BAC clones (1.3 million sequence reads). The shotgun sequence was compiled into an initial assembly (labeled Release 1.0) on December 11, 2003. This assembly was improved and released as version 1.1 in January 2004 and is now being analyzed by the insect genomics community. Assembly 1.1 includes 11,600 scaffolds covering a genome of 206 Mbp, with an N50 length of 223 kb. Forty percent of the assembly is covered by anchored and oriented scaffolds (n = 197). 94% of the 1,050 known honeybee markers are present in the assembly along with whole or partial sequences for 95% of genes described previously. Automated gene predictions using this and future assemblies will provide lists of known and de novo genes for honeybees. These predictions will be followed by efforts at Baylor and other institutions to manually annotate genes of interest to the medical and agricultural communities.

Assembly 1.1 used existing markers from known honeybee chromosomes to establish a physical map for bees, a product that will prove useful both for

placing genes of interest and for agricultural goals such as marker-assisted breeding. This map will evaluate the need to fill sequence gaps in critical areas, and will help predict the feasibility of marker or locus-specific breeding tools for honeybees. Future assemblies will be improved by the addition of BAC end sequences and other sequences generated to improve the sequence depth in targeted regions.

Funding for this project came from NHGRI ($6.9M) and USDA ($750K).

(Viewable on the web at: http://hgsc.bcm.tmc.edu/projects/honeybee/)

Chicken Genome Drafted at Washington University

The chicken is the first avian species for which whole genome sequence data is available in the public domain. When first proposed to NHGRI in 2002 (see original white paper proposal at: http://genome.wustl.edu/projects/chicken/Chicken_Genome.pdf), the chicken was identified as an extremely valuable species for genome sequencing due to: 1) its premier status as a primary model organism for studies in embryology and development, viruses and cancer, immunology, and gene regulation; 2) novel aspects of its biology relevant to human biology such as the Z/W sex determination system and high microchromosome content; 3) its intermediate evolutionary position between the mouse and fugu genomes; 4) its potential for providing a better understanding of the evolution of gene order and arrangement; 5) the wealth of genomics resources available from the chicken research community; and 6) its clear and growing importance as a source of animal protein in world agricultural production and the food system. Recent outbreaks of avian flu have accelerated scientists' interest in learning more about the chicken genome and how genetic variation may play a role in the susceptibility of different strains to the disease.

On March 1, 2004, NHGRI announced that a team led by Richard Wilson, Ph.D., from the Washington University School of Medicine in St. Louis successfully assembled the genome of the Red Jungle Fowl, *Gallus gallus*, which is the ancestor of domestic chickens. Sequencing of the chicken genome began in March 2003. NHGRI provided about $13 million in funding for the project.

The Washington University researchers have deposited the initial assembly, which is based on seven-fold sequence coverage of the chicken genome (approximately 1 billion bases), into GenBank (www.ncbi.nih.gov/Genbank). In turn, GenBank will distribute the sequence data to the European Molecular Biology Laboratory's Nucleotide Sequence Database, EMBL-Bank

(www.ebi.ac.uk/embl/index.html), and the DNA Data Bank of Japan, DDBJ (www.ddbj.nig.ac.jp).

To facilitate comparative genomic analysis, the researchers also have aligned the draft version of the chicken sequence with the human sequence. Those alignments can be scanned using the University of California, Santa Cruz's Genome Browser, (http://genome.ucsc.edu/cgi-bin/hgGateway); the National Center for Biotechnology Information's Map Viewer, (www.ncbi.nlm.nih.gov/mapview); and the European Bioinformatics Institute's Ensembl system, (www.ensembl.org).

In addition, using the *Gallus gallus* genome sequence assembled by Washington University as a reference framework, an international team, led by the Beijing Genomics Institute in China and supported by the Wellcome Trust in Great Britain, has created a map of genetic variation for three different strains of domestic chickens. The strains were a broiler strain from the United Kingdom, a layer strain from Sweden and a Silkie strain from China. To make the map, researchers identified and analyzed about 2 million genetic variation sites, mostly SNPs. The genetic variation data will soon be deposited into GenBank, from which the data will be freely accessible to researchers worldwide.

The project team has proposed to conduct a pre-finishing phase in 2004 to close gaps, extend the size of sequence contigs and enhance overall quality of the draft sequence. A crude analogy would be if the sequence is viewed as a book, pre-finishing takes the product from an unbound collection of thousands of loose-leaf pages to an ordered and bound manuscript. This process will provide a much more valuable and reliable product for end users and is very cost effective, as the clone libraries remain readily available for automated re-sorting. Funding for the pre-finishing project is under consideration by USDA.

For further information about the chicken genome project see (http://genome.wustl.edu/projects/chicken).

Broad Institute, MIT Well Underway in Sequencing the Dog Genome

Dogs are unique among mammalian species in the extent of variation they show in morphological traits such as height, weight, mass, shape, and behavior. Yet, within each breed, key traits are inherited within extremely narrow limits. No other mammalian species presents natural variation on such a scale, yet individuals from nearly any breeds can be mated to yield fertile offspring.

Given the aggressive breeding programs needed to reproducibly generate animals of distinctive size, shape and behavior, it is not surprising that purebred dog fanciers have also produced closed breeding populations, which carry over 400 inherited disorders. Genetic diseases are predicted to occur with high frequency in populations with closed gene pools and in which breeding of close relatives is used to propagate desired traits. Breeds established from a small number of founders and expanded rapidly to meet breeders' and consumers' demands suffer the most due to a combination of random drift and inbreeding depression. Autosomal recessive and complex traits present the biggest problem as the status of asymptotic carriers may not be suspected until several litters have been produced. This includes diseases such as cancer, heart disease, deafness, blindness, motor neuron disease, skin disorders, and a host of autoimmune disorders, each of which has been difficult to study in humans. The physiology, disease presentation and clinical response of dogs often mimic human diseases closely. Thus, positional cloning of genetic diseases in the dog can inform us about the genetic basis or susceptibility to similar diseases in humans.

Because of the potential of the dog as a model for studying human disease, the sequencing of the dog genome was approved by NHGRI in 2002 and is now nearing completion at the Broad Institute, MIT. A female Boxer named Tasha was chosen for sequencing because among the breeds examined by the Broad investigators the Boxer has the lowest amount of variation in its genome and therefore is likely to give an easily assembled genome sequence.

The sequencing of Tasha has now been completed to 6.5-fold coverage of the genome and sequence assembly is underway. The sequence traces are available at http://www.ncbi.nih/Traces/trace.cgi?. The assembly will be posted in late spring of 2004 and will be aligned against the human sequence. Automated annotation of the genome will also be provided. Tasha's sequence will then be compared to 100,000 sequence reads from 20 other breeds to generate approximately 25,000 SNPs to be used in genetic studies of the dog.

For more information about the dog genome project see: http://www.broad.mit.edu/media/2003/pr_03_tasha.html.

Unprecedented Partnership Launches the Bovine Genome Sequencing Project

On December 12, 2003, Secretary of Agriculture Ann Veneman and Francis Collins, Director of the National Human Genome Research Institute of NIH, jointly announced the launching of the International Bovine Genome Sequencing Project (BGSP). This milestone was reached after a monumental effort to build an international consortium to fund the $53M project.

In 2002, NHGRI identified the cow as a high priority through its proposal review process (white paper proposal can be viewed at http://www.genome.gov/10002154). At that time, NHGRI indicated that its support for the bovine genome project was dependent upon raising matching funds. Over the course of the following year, the IWG, with the assistance of the Alliance for Animal Genomics Research, worked to bring together an international consortium to provide the matching funds needed for the project. The International Bovine Genome Sequencing Consortium (IBGSC) was formed of the funding agencies that are contributing funds to match the total needed for the project. The group consists of (with funding contribution): National Institutes of Health, NHGRI ($25M), USDA ($11M), State of Texas ($10M), Genome Canada ($5M), Australia's CSIRO ($1M), Ag. Research New Zealand ($1M), and US Cattlemens Beef Board, Texas, and South Dakota Beef Councils ($820K). The project is underway at the Baylor College of Medicine's Human Genome Sequencing Center, Houston, TX

(BCM-HGSC), utilizing the international bovine BAC map as the scaffolding upon which to do the sequencing (see previous section of this report).

The goals of the project are to produce a draft sequence coverage, assembly and primary automated annotation of the bovine genome, including SNP discovery, full-length cDNA sequencing and finishing of a selected set of BACs. The genomic sequencing goal is 6-fold coverage (WGS and BAC skims combined) with an additional 1-fold coverage from WGS SNP discovery sequencing. This draft coverage will be augmented by finishing of 75 Mb of sequence from regions that contain genes influencing important traits. The cDNA sequencing will provide nearly 10,000 sequences.

The goal of Phase I will be to produce and assemble a draft bovine genome sequence, including a SNP discovery component and production of full-length cDNA sequences. The BCM-HGSC is the lead center in this phase and will coordinate the overall conduct of the project. The Genome British Columbia Sequencing and Mapping Platform (GBCSMP) at the British Columbia Cancer Agency (BCCA) will participate in this phase of the project through production of full-length bovine cDNAs. The annotation of the bovine genome sequence, although not funded through the IBGSC, is being carried out by the European Bioinformatics Institute (EBI) with funding from the Biotechnology and Biological Sciences Research Council (UK) in collaboration with the BCM-HGSC. Phase I of the project is targeted for completion in late 2005.

Project oversight is being performed by a Bovine Genome Sequencing Project Advisory Committee (BGSPAC). This committee is comprised of representatives from each contributing funding agency, a scientist chosen by that funding agency and a representative from each sequencing center. Ex-officio members include a representative of the Alliance for Animal Genomics Research, and a representative of the authors of the “white paper” proposal. Steven Kappes (MARC) and George Weinstock (BCM-HGSC) are co-chairs of this committee.

As of March 1, 2004, a total of 2,500,000 sequence reads had been deposited in the trace archive (amounting to ~0.5-fold sequence coverage). The sequence is available at http://www.ncbi.nig.gov/Traces/trace.cgi?.   Work on the BAC tiling path had also commenced with the construction of shotgun libraries. Efforts during the first three quarters of 2004 are being focused on completion of the whole genome shotgun portion of the project. Additionally, planning of the SNP discovery and full-length cDNA portions of the effort are underway by the BGSPAC.

Further information about the bovine genome project is available at: http://hgsc.bcm.tmc.edu/projects/bovine/.

International Swine Genome Sequencing Consortium Formed

Recent completion of the human genome sequence provides the starting point for understanding the genetic complexity of humans and how genetic variation contributes to diverse traits and disease. Additional species must be sequenced to resolve the genetic complexity of human evolution and to effectively use genetic information in veterinary and human medicine. The pig occupies an evolutionary position within mammals distinct from primates and rodents but close to the bovine and dog. Its importance in biomedical research and agriculture has led to the development of genomic resources that will be extremely important to the research community studying the pig. The pig has been widely used as a biomedical model for many years, and is expected to play a critical role in research directed toward understanding human obesity.

Consumption of pork continues to grow worldwide and pork represents 43% of all red meat eaten in the world. Given its importance in modern animal agriculture, obtaining the complete genome sequence of the pig is imperative. Such scientific information will lead to improved performance and production, decreased problems with porcine diseases and reduced environmental and welfare concerns.

Leaders from the pig genetics community met in France in September 2003 and established the International Swine Genome Sequencing Consortium (ISGSC) with representation from the US, UK, France, Denmark, Korea, Japan, and China. The goals of the ISGSC are to coordinate the scientific and funding activities associated with sequencing the pig genome. The scientific strategy includes: (1) contig sequencing (~2X coverage) and the assembly of all sequences at a major genome sequencing center and (2) whole genome shotgun sequencing (~5X coverage) that will be conducted at various centers around the world. All data will be rapidly placed in the public domain. Larry Schook from the University of Illinois is serving as chair of the consortium.

Similar to the approach being utilized in the bovine genome sequencing project, the international swine BAC map will be used as the scaffolding for the sequencing effort. The goal of the ISGSC is to secure the necessary funding required (approximately $40M) for the project by October 2004.

Commitments thus far to the project include release of the ~.7-fold whole genome shotgun sequence from the Sino-Danish pig project ($5M in kind), a pledge from Korea to contribute WGS at NLRI ($2M in kind), and release of BAC-

end sequence from the Sanger Center ($3M in kind), Roslin Institute ($1M in kind), Iowa State University ($0.3M), and the University of Illinois ($1M in kind). Meetings have been held with USDA, the U.S. pork industry (Pork Board and National Pork Producers Council), the European Commission, Genome Canada, the Wellcome Trust, BBSRC, DEFRA, INRA and the Danish, Canadian, and UK pork industries to explore potential direct funding of the project.

Genomics-Derived Tools Critical in US BSE (Mad Cow Disease) Crisis

When the first index case of bovine spongiform encephalopathy (BSE), more commonly referred to as "mad cow disease," was reported in Washington state in December of 2003, little did anyone know that it would become one of the first major cases for the application of genomics technology to providing a valuable and rapid solution in a crisis situation in animal agriculture. Following the identification of the diagnosis of this animal, U.S. beef market prices dropped precipitously, due to pressure imposed by the closure of most of the major export markets to U.S. beef.

Shortly after the identification of the BSE case, records associated with the particular cow in question indicated that she was originally imported from Alberta, Canada, where a similar case of BSE had been identified several months earlier. However, the records did not unequivocally prove her origin. Fortunately, the records identified the sire and other relatives of the BSE index case; furthermore, their tissues were available for genetic analyses. The availability of these tissue samples permitted rapid pedigree verification, using genomic technology, and thus confirmed the records indicating the Canadian origin of the BSE index cow.

The technology used in this case was developed in 2002 by ARS scientists Mike Heaton and coworkers at the USDA/ARS U.S. Meat Animal Research Center, Clay Center, Nebraska. They identified 32 selected single nucleotide polymorphisms (SNPs) in cattle that provided very accurate identification of individual animals. This SNP battery was derived from hundreds of SNPs and chosen for maximum allelic diversity. Based upon results from multiple breeds, this genotyping system is capable of discriminating between any two random, unrelated individuals in the population at a probability level of $2x10^{-13}$ (i.e. odds ratio of 1:1 trillion).

Using this genotyping test, ARS scientists Will Laegreid, Mike Clawson, and Mike Heaton were able to provide strong scientific evidence that the Washington

state index cow was one of a group of 81 females that had crossed the U.S. border from Canada several years earlier. This provided the missing link in the traceability of this female to verify that she was born prior to the 1997 implementation of a full ban on animal protein feeding in Canada.

A month before the Washington state BSE case was identified, ARS scientists reported the complete prion gene sequence variation in U.S. cattle. A quick comparison of results showed that the BSE index cases did not have novel prion gene sequences. This was important because DNA sequence analysis of both the May 20th, 2003, Canadian BSE case and the Washington state case had revealed apparently novel, but different, mutations in their prion gene sequences. This would have otherwise suggested the occurrence of "sporadic BSE," i.e. not related to contaminated feed. Together, the new bovine genomic tools had an immediate and positive impact on the beef industry, DNA traceback, and animal disease control.

V. Plans for FY 2004

Beyond Sequencing–Focusing on Bioinformatics and Functional Genomics Tool Development for Domestic Animal Species

With the establishment of a genomics infrastructure for a number of the domestic animal species, the Interagency Working Group is now turning its attention to its goals in the areas of bioinformatics and functional genomics. To reiterate, those specific goals are:

- Data management and bioinformatics to specifically support agriculturally important species. This will support significant improvements in data management and analysis software, allow greater data accessibility and secure long term maintenance, increase capabilities to deal with rapidly accumulating data complexity as databases include functional information, and provide more powerful tools to mine large genomes (individually and comparatively).
- Functional genomics to specifically investigate agriculturally important species. An increase in data for livestock genomes requires concomitant investment in functional genomics to support genome annotation, the study of gene regulation and expression, and species evolutionary relationships. Researchers will have access to individual genes, which can be cloned and characterized, the ability to scan entire genomes for specific functions, and be able to gain new insights from comparisons to other genomes. Quantitative trait loci will be accessible on a gene level, leading to studies that will elucidate characteristics such as food productivity and disease

resistance and will lead to the ability to manipulate those traits more quickly and effectively.

USDA is planning a workshop be held in Washington, DC on September 22-23, 2004 to specifically develop a road map for addressing the infrastructure needs in domestic animal bioinfomatics and functional genomics. This workshop will bring together leaders from the domestic animal genomics research community with leading authorities from the human and model species communities to flesh out this plan. The resulting white paper report will be presented to the IWG for consideration in carrying out its mission to move efforts forward in these areas.

Development of Broad-Based Functional Genomics Programs in Domestic Animal Infectious Disease

One of the most important potential applications of domestic animal genomics is in the area of infectious animal disease. Recent attention on disease outbreaks with potential human impact, such as the avian flu outbreak in Asia, North America, and Europe, and prion diseases such as BSE in cattle, scrapie in sheep and chronic wasting disease in elk and deer, have accentuated the need for better understanding of the genomes of both host and pathogen. Genomics offers the opportunity to better characterize genetic susceptibility to threatening diseases, particularly those transmitted between humans and animals. With the completion of annotated genome sequences for the major domestic animal species, the opportunities to pursue these genomics approaches are at hand.

In the coming year, the question of what will be required for the publicly funded research community to develop broad-based domestic animal genomics research programs to address these critically important animal disease and biosecurity issues needs to be addressed. This is related to the Presidential Directive HSPD-9 which calls for a national plan for protection of agricultural interests within Homeland Security.

Expanding the Genomics Infrastructure to Include Additional Species

Expanding the genomics infrastructure beyond the major domestic animal species focused upon in the IWG's initial efforts (chicken, cow, pig, dog, honeybee) is a major goal in the coming year. Efforts are underway in a number of other species to develop community resources in the public domain

including BAC maps and EST collections as well as draft genome sequence assemblies.

The Department of Energy's Joint Genome Institute at Walnut Creek, California recently announced that it will accept white paper proposals for whole genome shotgun sequencing through their Community Sequencing Program. A number of species of interest to the IWG were submitted including the channel catfish, rainbow trout, and Pacific oyster.

While not a domestic animal, a mammal of considerable interest to marine mammologists is the dolphin. Efforts in the coming year by the Office of Naval Research are seeking to launch a dolphin genomics effort.

The sheep genomics community has recently launched an effort to build a comprehensive BAC map of the sheep genome, working with the Children's Hospital of Oakland, Oakland, California. The domestic cat also continues to be rated at a "Moderate" priority for sequencing by NHGRI.

Does Genomics Technology Offer the Solution for Animal Traceability and a National Animal ID System?

As described earlier in this report, DNA genotyping was used in the recent U.S. BSE crisis through its use to verify the Canadian origin of the Washington state BSE index cow. Almost immediately following the discovery of this case, the Secretary of Agriculture directed the establishment of a national animal ID system to allow full traceability of animals in such biosecurity events in the future. Planning for a national animal ID system to allow traceability was already well underway, however.

In October of 2003, the U.S. Animal Identification Plan (USAIP), developed under the facilitation of USDA's Animal and Plant Health Inspection Service, was endorsed by the U.S. Animal Health Association. While the USAIP lays out the basic plan for a national animal identification system, with the goal of allowing traceability of an animal within 48 hours of a biosecurity breach, it does not stipulate the platform for the system. DNA genotyping has high potential for providing such a platform.

By increasing the genotyping system used by USDA to verify the parentage of the BSE index cow in 2003 from 32 to 45 SNPs, the discrimination power can be raised to a level that allows the unique identification of any animal drawn randomly from the cattle population, although allelic frequencies of additional

SNPs must be validated across additional industry breed populations. Currently, an industry collaborator has moved the SNP battery to include 47 SNPs (38 highly informative SNPs, 7 Brahman-specific (i.e. Indicus) SNPs and two sex-determining SNPs).

The application of genomic technology to livestock improvement and biosecurity is beginning to occur. Several DNA diagnostic tests are now commercially available in the public domain with a number under industry developmental testing. An even greater number of markers are routinely being used within private industry, particularly in the case of swine. The number of DNA diagnostics is expected to increase exponentially with the release of annotated whole genome sequence data for chicken, cattle, and swine within the next few years, the increase in technologies allowing clearer analysis of gene function and regulation (microarrays, serial analysis of gene expression, RNAi, etc), and the integration of genetic and physical maps for these species (integration of linkage, radiation hybrid, and BAC maps). Such DNA diagnostics will allow the determination of DNA genotypes of individuals for both genetic improvement and precision management applications, including prevention of risk of disease outbreak.

The national animal ID system could be the vehicle to allow the widespread use and adoption of DNA-based management systems. The determination of a unique animal ID, based upon the animal's DNA profile for a pre-determined SNP battery, could serve two purposes. First, unique identification of an animal is easily accomplished. Secondly, the SNP battery can be designed to simultaneously genotype economically important diagnostics for genetic and management purposes for those who wish to have access to such information.

Previous use of DNA genotyping for animal identification has been almost exclusively for parentage verification. Widespread implementation of the technology has been very slow, primarily due to the high cost ($25 per sample), in using microsatellite-based genotyping systems. Economic modeling in beef cattle systems has shown that the cost of this technology must be in the $5 to $7 per head range to be practical. Higher volume would be expected to substantially reduce per head cost.

Current genotyping platforms have achieved remarkable throughput and costs have lowered dramatically in recent years. Current estimates range from $.25 to $.50 per SNP in high volume operations. Using a mass spectrometry platform, most companies estimate that the lower end of this cost range is feasible. Thus, a DNA genotyping system employing 40 markers could be commercially

available for $10 or less in high volume systems, as would be the case in a national animal ID scheme. Several commercial vendors are attempting to enter various parts of this type of market.

The practical application of a DNA-based system would operate in the following manner:

- An animal would be sampled for DNA simultaneously when a unique bar code is assigned to it as its individual animal ID (USAIN in the USAIP). Tagging systems allowing such to occur using electronic ID (EID) are available in the marketplace which allow retrieval of DNA from blood spotted to a card at the time of tagging, already cross-referenced to the bar code of the tag.
- The DNA genotype for the animal would be determined and entered in the national database by cross-reference to the bar code.
- If the animal was involved in a situation requiring trace-back and verification, DNA would always be available, through to the retail cut basis, to do so.
- If external identification for an animal was lost, the animal could be easily re-genotyped to determine its original USAIN.
- If the DNA genotype information was desired by the owner of the animal, the commercial vendor could make such available on a fee basis, allowing the offsetting of a portion of the genotyping cost of the national animal ID system.

The application of genomics-based technology in a national animal identification system will be carefully evaluated within federal research programs in the coming year.

VI. Appendix

URLs for useful information on domestic animal genomics:

National Science and Technology Council homepage:
http://www.ostp.gov/NSTC/html/NSTC_Home.html

USDA Research, Education, and Economics homepage:
http://www.reeusda.gov/ree/

NSF Directorate for Biological Sciences homepage:
http://www.nsf.gov/bio

DOE Office of Science homepage: http://www.sc.doe.gov/

National Human Genome Research Institute homepage:
http://www.nhgri.nih.gov/

DbEST (EST database at the NCBI): http://www.ncbi.nlm.nih.gov/dbEST/

Online Mendelian Inheritance in Man Database at NCBI:
http://www.ncbi.nlm.nih.gov/Omim

Genbank at NCBI: http://ww.ncbi.nih.gov/Genbank

National Center for Biotechnology Information's Map Viewer:
http://ww.ncbi.nlm.nih.gov/mapview

European Molecular Biology Laboratory's Nucleotide Sequence Database, EMBL-Bank:
http://ww.ebi.ac.uk/embl/index.html

DNA Data Bank of Japan, DDBJ: http://ww.ddbj.nig.ac.jp

University of California, Santa Cruz's Genome Browser:
http://genome.ucsc.edu/cgi-bin/hgGateway

European Bioinformatics Institute's Ensembl: http://ww.ensembl.org

Washington University Genome Sequencing Center:
http://genome.wustl.edu/projects

Broad Institute, MIT: http://www.broad.mit.edu/

Baylor College of Medicine Human Genome Sequencing Center:
http://hgsc.bcm.tmc.edu

ChickNet: http://chicken-genome.org

Alliance for Animal Genomics Research:
http://eversoleassociates.com/alliance

Center for Integrated Animal Genomics, Iowa State University:
http://ciag.iastate.edu

Biotech Center, University of Illinois: http://www.biotec.uiuc.edu

USDA Agricultural Research Service: http://www.ars.usda.gov

USDA/ARS U.S. Meat Animal Research Center:
http://www.sol.marc.usda.gov/

USDA/ARS Beltsville Animal and Natural Resources Institute:
http://lpsi.barc.usda.gov/

Acknowledgements

The Interagency Working Group on Domestic Animal Genomics acknowledges the assistance of the following individuals in the preparation of this report:

Jane Peterson, NIH/NHGRI
Deb Hamernik and Muquarrab Qureshi, USDA/CSREES
Kevin Hackett, Steve Kappes, Gary Rohrer, and Warren Snelling, USDA/ARS
Larry Schook, University of Illinois
Avital BarShalom, OSTP (previous IWG Executive Secretary)
Ruth Coy, Jody Shuart, and Mina Chung, USDA/ARS Information Staff

Photo Credits

Cover	Left to right: Piglets, photo by Scott Bauer, ARS image number K9455-1; bee, photo by Jack Dykinga, ARS image number K5400-1; chicks, photo by Keith Weller, ARS image number K3627-16; Jack Russell terrier, photo by Dave Hodge; and sheep, ARS image number K4166-5.
Page vi	Sheep graze near Odell Lake at the U.S. Sheep Experiment Station in the Centennial Mountains of southwestern Montana, photo by Scott Bauer, ARS image number K5629-2.
Page 1	Jack Russell terrier, photo by Dave Hodge.
Page 2	Chicks atop a picture of a genetic map of a chicken, photo by Peggy Greb, ARS image number K8764-1.
Page 3	Cat/Human Parallel Radiation Hybrid Map. National Cancer Institute.
Page 4	Entomologist Steve Sheppard prepares an agarose gel to be used in separating honey bee DNA fragments, photo by Scott Bauer, ARS image number K5799-1.
Page 5	Differences between European and Africanized honey bees can be seen in this DNA sequencing gel being read by microbiologist Hachiro Shimanuki and geneticist Cristina Arias, photo by Scott Bauer, ARS image number K5764-16.
Page 6	Cattle feeding, photo by Brian Prechtel, ARS image number K5643-20.
Page 7	A small dairy farm in western Maryland, ARS image number K8502-1.
Page 8	Calf, photo by Scott Bauer, ARS image number K9481-1.
Page 9	Goat, photo by Scott Bauer, ARS image number K8595-9.
Page 10	Cat on human conserved synteny map, National Cancer Institute.

Page 11 White turkey, photo by Scott Bauer, ARS image number K7043-16.

Page 12 Chicks, photo by Keith Weller, ARS image number K3627-16.

Page 14 Bee, photo by Jack Dykinga, ARS image number K5400-1.

Page 15 Red jungle fowl, photo by Bill Payne.

Page 16 Boxer, Whitehead Institute/MIT Center for Genome Research.

Page 17 Sheep, ARS image number K4166-5.

Page 18 USDA ceremony launching the bovine genome sequencing project, Washington, DC, December 12, 2003. Left to right: Joseph Jen, Under Secretary, REE, USDA; Wayne Roberts, Budget Director, State of Texas; Hon. Ann Veneman, Secretary of Agriculture, USDA; Francis Collins, Director, Human Genome Research Institute, NIH; Kathie Olsen, Associate Director, OSTP; and Martin Godbout, President and CEO, Genome Canada. Photo by USDA Communications Office.

Page 19 Hereford bull, photo by Mike MacNeil.

Page 20 Piglets, photo by Scott Bauer, ARS image number K9455-1.

Page 21 Cattle, photo by Keith Weller, ARS image number K4328-8.

Page 22 Lamb, photo by Karen Carpenter, Texas A&M University.

Page 23 St. Croix sheep, ARS image number K3719-1.

Page 24 Siamese cat, photo by Ann Carlson.

Page 25 Dairy cow, photo by Keith Weller, ARS image number K5176-3.

Executive Office of the President

Office of Science and Technology Policy

Washington, D.C. 20502